



United States
General Accounting Office
Washington, D.C. 20548

General Government Division

B-249150

FEDERAL DATA COLLECTION: Agencies'
Use of Consistent Race and Ethnic
Definitions

December 15, 1992

The Honorable John Conyers, Jr.
Chairman, Committee on Government Operations
House of Representatives

Dear Mr. Chairman:

This report responds to your request that we review federal agencies' definitions of race and ethnicity as applied to data collection and, if noted, determine causes for inconsistent reporting of such data. You were concerned about the comparability of data that agencies collect because you had found that the Justice Department did not use common definitions of race and ethnicity in their reporting of the population at various stages of the criminal justice system.

Results in Brief

We found that federal agencies are required to use consistent definitions of race and ethnicity in the development of their own data collection methodology. This requirement is to be enforced by an Office of Management and Budget (OMB) control process. This process includes a governmentwide policy that provides specific definitions for race and ethnicity and a mandatory review of agency-proposed surveys by OMB to determine compliance with the policy. In addition, federal data collection agencies are to use consistent definitions when relying on the Census Bureau for data because the Bureau is required to adhere to the OMB policy as well.

Inconsistent use of racial and ethnic terms can occur, however, in the reporting of data when agencies use external sources such as state-provided data. Although federal data collectors provide states with race and ethnic data definitions based on the OMB policy, many states do not record their populations by the same criteria. For example, one state records only Puerto Ricans as ethnic Hispanic, while others record all Spanish-speaking peoples in this category. Differences in state data collection from the federal policy are sometimes based on the particular racial and ethnic make-up of a state's population. This situation has caused some federal data collection agencies to make their own surveys in order to obtain more reasonably accurate projections of racial and ethnic populations.

Inconsistent use of racial and ethnic terms can also occur when people are classified by observer-identification instead of self-identification. Agencies

we contacted said that self-identification is the preferred method for categorizing individuals by race or ethnicity for most agency surveys because observer-identification can result in misidentifications. However, when agencies use data that are collected by nonfederal sources, self-identification may not have been used.

Background

The enactment of the Civil Rights Act of 1964 (PL 88-352) created the need for the federal government to collect statistics on race for compliance and enforcement purposes in such areas as education and housing. There were no standard definitions or procedures for collecting such data in the early 1960s, but a need for both became apparent by the 1970s as a variety of public and private entities began using racial data collected by various federal agencies for policy and trend analysis. A desire for data on the status of an emerging multiracial Hispanic population created a need for data by ethnic group as well.

In 1973, an interagency committee studying the educational status of minorities found, among other things, a deplorable lack of useful data on racial and ethnic groups. This discovery led to the establishment of a task force, consisting of staff members of the Department of Health, Education, and Welfare (HEW) and interested agencies that collected or used racial/ethnic data, to explore the topic. This effort resulted in the development of a governmentwide policy for federal agencies to use when collecting, reporting, and maintaining data on race and ethnicity. The policy eventually evolved into OMB Statistical Policy Directive No.15 (see app. I). This directive also responded to concerns that both the executive branch and Congress expressed about the collection and use of compatible, nonduplicated, exchangeable racial and ethnic data by federal agencies. The directive also was part of the Paperwork Reduction Act changes in federal information management.

Directive No.15, which has been in effect since 1980, provides standard classifications for the recordkeeping, collection, and presentation of data on race and ethnicity in civil rights compliance, general program administrative and grant reporting, and statistical reporting. It requires all agencies to comply with minimum statistical criteria and defines four racial groups (American Indian or Alaskan Native, Asian or Pacific Islander, Black, and White) and one ethnic group (Hispanic), based on geographical or cultural, rather than scientific, distinctions. Although agencies are allowed to collect more detailed information, they must aggregate the data to these basic categories. The preferred method for

categorizing individuals (self- vs. observer-identification) is not stated. In the case of persons who are of mixed racial and/or ethnic origins, the directive recommends using the category that most closely reflects the individuals' recognition in their community.

Objectives, Scope, and Methodology

As agreed with the Committee, the objectives of our review were to examine federal agencies' definitions of race and ethnicity for collecting data on populations and, if noted, determine causes for inconsistency in the reporting of these data.

In carrying out our work, we reviewed OMB statistical policies that provide rules and guidance to federal data collectors, interviewed agency officials responsible for data collection methodology, reviewed literature and research papers on the subject, and reviewed the American Statistics Index (ASI)—the most complete compilation of federal statistical reports.

We identified eight federal statistics agencies that, according to OMB, collect significant amounts of racial and ethnic data and receive data from state and local sources. These agencies also collect and publish reports on significant portions of the U.S. population and publish the most reports on conditions of the population by race and ethnicity, according to ASI (see app. II). In addition to the Justice Department, we selected three of the agencies (the Department of Education, the Department of Health and Human Services, and the Bureau of Labor Statistics) that OMB indicated might have been affected the most by nonfederal data providers such as states. We interviewed the officials who were responsible for data collection methodology at each of the four agencies. We also obtained and reviewed a sampling of survey/questionnaire documents, data collector and respondent instructions, and reports based on collection efforts. The sample documents we obtained were typical of the collection instruments supporting the agencies' major data systems. We then interviewed the remaining four agencies—the Department of Housing and Urban Development (HUD), the Small Business Administration, the Equal Employment Opportunity Commission, and the Department of Defense—to determine their methods for collecting race and ethnic identified data.

We did not evaluate the adequacy, appropriateness, or validity of the federal definitions for race or ethnicity. Our work was done between December 1991 and August 1992 in accordance with generally accepted government audit standards.

Agencies Used Standard Definitions in Data Collection

The eight agencies we reviewed used the standard definitions of Directive No. 15 in the data collection operations we examined. We found that these standard definitions were contained in data collection instruments, instructions, and survey documents and also were published in the resulting reports, usually in the methodology statement.

The consistent use of definitions is accomplished by an OMB control procedure that is designed to help assure that standards are properly incorporated in data collection efforts. The Office of Information and Regulatory Affairs, the OMB unit that is responsible for establishing the race/ethnic definition policy, is required to approve all federal data collection instruments and methodologies through the Paperwork Reduction Act regulations, before an agency begins collecting data.¹ The limitation of this control process is that it monitors only the use of the standards in the development of the methodology, not the results of the collected data. OMB officials explained that they do not routinely verify agency compliance with the policy beyond survey methodology approval.

Another practice that helps assure that the policy's definitions are followed is the federal statistical agencies' extensive use of Census Bureau support and products that also are governed by the rules of Directive No. 15. For example, the Bureau of Justice Statistics, the Bureau of Labor Statistics, the National Center for Education Statistics, the National Center for Health Statistics, HUD, and the Defense Manpower Data Center use elements of the Decennial Census or the monthly Current Population Surveys in their reports. Furthermore, some of the agencies use the Census Bureau as a field agent for data collection because the Bureau's survey capability goes beyond agency capabilities. In both situations, the agencies use data that are collected using methodologies that OMB has approved.

Sources of Inconsistent Reporting of Racial/Ethnic Data

Although federal data collection agencies are required to follow Directive No. 15 in developing their collection methodologies, we found some consistency problems in the agencies' reporting when they used data not controlled by federal collection rules or data based on observer-identification. Reliance on data that are collected by state or local sources has created consistency problems in Justice and Education statistical reports. The other agencies we reviewed did not use racial data from external sources in their reports. Observer-identification can lead to

¹OMB can change an agency's data collection scope through this process as well. For example, in 1986 OMB did not approve the Federal Bureau of Investigation's continued collection of ethnic data on arrestees, which had the effect of discontinuing data collection on Hispanic arrests.

incorrect classification of individuals, as a recent Centers for Disease Control (CDC) study disclosed. In that study, birth and death statistics were compared for infant mortality over a 3-year period. A large portion of minority infants were incorrectly classified as White because recorders were making racial and ethnic judgments based on observation.²

Problems Created by State-Provided Data

The Justice Department produces a wide variety of statistical reports on crime, the justice and correctional processes, and the criminal population. When Justice controls data collection, it is required to follow the OMB policy standards for race and ethnicity. However, to produce more geographically detailed reports, Justice depends on state or local data input, which can introduce error because of the various ways states categorize their populations.

Although federal agencies are permitted to use state- or local-generated data, the management of these data is outside of their jurisdiction. However, federal data collectors do provide these sources with racial and ethnic definitions and ask that they note any exceptions to definitions on information provided to the agency. When using these data in their reports, federal agencies may note these exceptions as well. A 1991 Justice Department analysis of prisoner population data from state sources disclosed about a 7 percentage point difference in categorization by race. Some states report Asians and American Indians in the White category, while others report their race as unknown. Some states also include Hispanics in the White category, regardless of race, while others consider Hispanics of unknown race. These situations occur in prisoner, parole, and probation population reporting by states and make accurate analysis of race and ethnic composition of the correctional system population difficult.

The different ways in which states categorize by race and ethnicity affect analysis of the Hispanic population in various Justice groups. For example, 19 states did not report the ethnic composition of their probation populations; 3 of these states (Illinois, California, and New Jersey) are among those in which the greatest concentration of Hispanics reside. Therefore, relying on state-furnished data does not provide an accurate picture of Hispanics in the corrections system as shown in table 1. Based on such state-reported data, Hispanic populations in these correctional

²Robert A. Hahn, Joseph Mulinare, Steven M. Teutsch, "Inconsistencies in Coding of Race and Ethnicity Between Birth and Death in U.S. Infants," The Journal of the American Medical Association (Jan. 8, 1992), pp. 259-263.

activities could be significantly understated because of the large percentage of the population for which ethnicity is unknown.

Table 1: State-Furnished Data on Hispanics in the Justice System

Correctional activity	Total population	Percentage identified as Hispanic	Ethnicity unknown ^a
Probation	2,461,333	5.8%	54.2%
Prisons	653,392	11.4	36.5
Parole	435,385	13.1	37.8
Arrests	14,340,900	^b	^b

^aThese figures represent the percentage of total population where states either did not report ethnicity or ethnicity was reported as unknown.

^bAs noted in footnote 1, Justice discontinued collecting data on Hispanic arrests in 1986. According to Justice officials, Hispanic arrest information will be collected in the future after a new crime incident reporting system becomes fully operational.

Source: Bureau of Justice Statistics, state-furnished data.

Although the Justice Department provides states with definitions for the race and ethnic data requested, it also asks the states to note deviations from the federal standards. Justice believes that states are reliable about reporting deviations and indicates the exceptions when using the data. Justice further believes that the state data are useful because they add detail to national statistics, cost the federal government nothing, and, in the case of arrest data, provide valuable real-time information recorded at the crime scene. Justice believes that the fact that it makes periodic projectable surveys that collect data on characteristics—including race and ethnicity—of crime-related populations (prisoners, victims, arrestees, parolees, etc.), compensates for unreliable state data. For example, every 5 years, Justice samples and interviews about 14,000 state and 8,500 federal inmates to obtain basic data on prison population characteristics. Using this procedure, Justice reported in 1991 that states undercounted their Asian and American Indian prisoner populations by 2 percentage points. Justice also found in this same evaluation that states undercounted Hispanics by about 5 percentage points. However, only prison and jail populations are now surveyed on a regular basis; parolees, those on probation, and arrestees are not. Justice explained that it has not had the resources to do these other surveys.

State reporting can also be incomplete. The National Center for Education Statistics (NCES) frequently surveys states for student population. Surveys

are needed to update the Center's Common Core of Data (CCD), which is a principal source for educational analysis. Although NCES officials believe that NCES receives good counts for the student population, racial and ethnic characteristics are sometimes incomplete. In a recent report (school year 1989-90), 9 out of 56 states and locations did not report the racial composition of their public school populations, which precludes national aggregation of student population by race.

State-furnished education data can also have a certain inherent inconsistency when race or ethnicity must be determined for students of mixed parentage. According to Education officials, some states determine a student's race or ethnicity by that of the mother, whereas others use the father's race or ethnicity. Because OMB guidance states that race or ethnicity for this situation should be that "which most closely reflects the individual's recognition in his community," the label applied by state policy may not be the same. Education officials explain that because they lack resources, they do not verify the conditions under which statistical information is gathered by the states and, consequently, have no idea of the extent of this problem.

The categorization of multiracial/ethnic peoples also is a problem for a federal agency if it uses data from states that define this group differently than the OMB policy dictates. A recent survey of 800 school districts by Education's Office for Civil Rights found that about 30 percent of them use a special category to classify people of mixed race or ethnicity. This category differs from the OMB policy that states that a person from this background should be categorized in one of the four racial or one ethnic categories that the individual is viewed by his or her community. This difference between federal and state policy could become a greater problem in the future because the number of multiracial and ethnic families has grown significantly over the last 20 years.

Problems Created by Observer-Identification

Another area of potential inconsistency occurs when others determine an individual's race or ethnicity. CDC reported in January 1992 that infant race was reported incorrectly on 3.7 percent of death certificates of all U.S. infants who died from 1983 to 1985. CDC's study of vital statistics databases of birth and death certificates for infants disclosed that 43.2 percent of Asian and American Indians were classified by race differently at death than at birth. These infant deaths were usually classified as White, thereby overstating White mortality somewhat but greatly understating Asian and American Indian infant mortality. CDC officials believe that death

certificate preparers were classifying infants by observation and not by following the proper procedure of asking a family member to identify the appropriate race or ethnic background of the infant.

Agency officials we interviewed generally believe that self-identification is more appropriate because of the problem of misidentification, which can occur when people are classified by observation. This situation is especially true as multiracial and multiethnic groups increase in relative size. However, the race identification issue is sensitive. In 1988, OMB proposed to amend its policy specifying only self-identification as the acceptable method for identifying race or ethnicity. OMB officials said that differing institutional and public comments on the draft policy amendment as well as the lack of consensus from public comment led to the initiative being dropped.³

Conclusions

Race and ethnic background is a sensitive subject to individuals and making determinations of such background for statistical purposes is an imprecise effort at best. The federal policy that addresses this area, Directive No. 15, attempts to provide consistency for federal data collection. Although the agencies we reviewed appear to follow the rules of Directive No. 15, they do not verify compliance with the definitions by data collectors during surveys because of the great demands on resources such an effort would require. This failure to verify compliance, combined with OMB's limited review, could be a source of error in reporting as shown by the CDC infant mortality study.

In addition, when agencies use data that are not collected under federal standards, inconsistency can be introduced because these sources may have their own rules that may differ from federal standards. On the basis of the reports we examined, the agencies we reviewed appeared to understand this condition and took it into account when publishing reports.

Agency Comments

We discussed the contents of this report with responsible OMB, Justice Department, and Education Department officials, who generally agreed with its accuracy. Their comments have been incorporated in this report where appropriate.

³Our review of academic literature related to OMB policy shows institutions, organizations, and individuals were concerned that the definitions are not accurate or complete, but no consensus exists on a better alternative or policy.

As arranged with the Committee, we plan no further distribution of this report until 30 days after the date of issuance, unless you publicly announce its contents earlier. At that time, we will send copies of this report to the Commissioner, National Center for Education Statistics; the Director, National Center for Health Statistics; the Director, Bureau of Justice Statistics; and the Director of OMB. Copies will be made available to other interested parties upon request.

The major contributors to this report were Xavier Richardson, Assistant Director; John Van Lonkhuyzen, Evaluator-in-Charge; and Kiki Theodoropoulos, Writer-Editor. If you have any questions about this report please contact me on (202) 275-8676.

Sincerely yours,

A handwritten signature in black ink, reading "L. Nye Stevens". The signature is fluid and cursive, with the first letters of the first and last names being capitalized and prominent.

L. Nye Stevens
Director, Government Business
Operations and Information Issues

OMB's Policy for Race and Ethnic Definitions

DIRECTIVE NO. 15¹

RACE AND ETHNIC STANDARDS FOR FEDERAL STATISTICS AND ADMINISTRATIVE REPORTING

This Directive provides standard classifications for recordkeeping, collection, and presentation of data on race and ethnicity in Federal program administrative reporting and statistical activities. These classifications should not be interpreted as being scientific or anthropological in nature, nor should they be viewed as determinants of eligibility for participation in any Federal program. They have been developed in response to needs expressed by both the executive branch and the Congress to provide for the collection and use of compatible, nonduplicated, exchangeable racial and ethnic data by Federal agencies.

1. Definitions

The basic racial and ethnic categories for Federal statistics and program administrative reporting are defined as follows:

a. *American Indian or Alaskan Native*. A person having origins in any of the original peoples of North America, and who maintains cultural identification through tribal affiliation or community recognition.

b. *Asian or Pacific Islander*. A person having origins in any of the original peoples of the Far East, Southeast Asia, the Indian subcontinent, or the Pacific Islands. This area includes, for example, China, India, Japan, Korea, the Philippine Islands, and Samoa.

c. *Black*. A person having origins in any of the black racial groups of Africa.

d. *Hispanic*. A person of Mexican, Puerto Rican, Cuban, Central or South American or other Spanish culture or origin, regardless of race.

¹ Directive No. 15 supersedes section 7(h) and Exhibit F of OMB Circular No. A-46 dated May 3, 1974 and as revised May 12, 1977.

e. *White*. A person having origins in any of the original peoples of Europe, North Africa, or the Middle East.

2. Utilization for Recordkeeping and Reporting

To provide flexibility, it is preferable to collect data on race and ethnicity separately. If separate race and ethnic categories are used, the minimum designations are:

a. Race:

- American Indian or Alaskan Native
- Asian or Pacific Islander
- Black
- White

b. Ethnicity:

- Hispanic origin
- Not of Hispanic origin

When race and ethnicity are collected separately, the number of White and Black persons who are Hispanic must be identifiable, and capable of being reported in that category.

If a combined format is used to collect racial and ethnic data, the minimum acceptable categories are:

- American Indian or Alaskan Native
- Asian or Pacific Islander
- Black, not of Hispanic origin
- Hispanic
- White, not of Hispanic origin.

The category which most closely reflects the individual's recognition in his community should be used for purposes of reporting on persons who are of mixed racial and/or ethnic origins.

In no case should the provisions of this Directive be construed to limit the collection of data to the categories described above. However, any reporting required which uses more

Appendix I
OMB's Policy for Race and Ethnic
Definitions

detail shall be organized in such a way that the additional categories can be aggregated into these basic racial/ethnic categories.

The minimum standard collection categories shall be utilized for reporting as follows:

a. *Civil rights compliance reporting.* The categories specified above will be used by all agencies in either the separate or combined format for civil rights compliance reporting and equal employment reporting for both the public and private sectors and for all levels of government. Any variation requiring less detailed data or data which cannot be aggregated into the basic categories will have to be specifically approved by the Office of Federal Statistical Policy and Standards for executive agencies. More detailed reporting which can be aggregated to the basic categories may be used at the agencies' discretion.

b. *General program administrative and grant reporting.* Whenever an agency subject to this Directive issues new or revised administrative reporting or recordkeeping requirements which include racial or ethnic data, the agency will use the race/ethnic categories described above. A variance can be specifically requested from the Office of Federal Statistical Policy and Standards, but such a variance will be granted only if the agency can demonstrate that it is not reasonable for the primary reporter to determine the racial or ethnic background in terms of the specified categories, and that such determination is not critical to the administration of the program in question, or if the specific program is directed to only one or a limited number of race/ethnic groups, e.g., Indian tribal activities.

c. *Statistical reporting.* The categories described in this Directive will be used as a minimum for federally sponsored statistical data collection where race and/or ethnicity is required, except when: the collection involves a sample of such size that the data on the smaller categories would be unreliable, or when the collection effort focuses on a specific racial or ethnic group. A repetitive survey shall be deemed to have an adequate sample size if the racial and ethnic data can be reliably aggregated on a biennial basis. Any other variation will have to be specifically authorized by OMB through the reports clearance process (see OMB Circular No. A-40). In those cases where the data collection is not subject to the reports

clearance process, a direct request for a variance should be made to the OFSPS.

3. Effective Date

The provisions of this Directive are effective immediately for all *new* and *revised* recordkeeping or reporting requirements containing racial and/or ethnic information. All *existing* recordkeeping or reporting requirements shall be made consistent with this Directive at the time they are submitted for extension, or not later than January 1, 1980.

4. Presentation of Race/Ethnic Data

Displays of racial and ethnic compliance and statistical data will use the category designations listed above. The designation "nonwhite" is not acceptable for use in the presentation of Federal Government data. It is not to be used in any publication of compliance or statistical data or in the text of any compliance or statistical report.

In cases where the above designations are considered inappropriate for presentation of statistical data on particular programs or for particular regional areas, the sponsoring agency may use:

(1) The designations "Black and Other Races" or "All Other Races," as collective descriptions of minority races when the most summary distinction between the majority and minority races is appropriate;

(2) The designations "White," "Black," and "All Other Races" when the distinction among the majority race, the principal minority race and other races is appropriate; or

(3) The designation of a particular minority race or races, and the inclusion of "Whites" with "All Other Races," if such a collective description is appropriate.

In displaying detailed information which represents a combination of race and ethnicity, the description of the data being displayed must clearly indicate that both bases of classification are being used.

When the primary focus of a statistical report is on two or more specific identifiable groups in the population, one or more of which is racial or ethnic, it is acceptable to display data for each of the particular groups separately and to describe data relating to the remainder of the population by an appropriate collective description.

Federal Agencies Contacted During This Review

Department of Justice (Bureau of Justice Statistics)

Department of Education (National Center for Education Statistics and
Office for Civil Rights)

Department of Health and Human Services (National Center for Health
Statistics)

Department of Labor (Bureau of Labor Statistics)

Department of Defense (Defense Manpower Data Center)

Department of Housing and Urban Development

Equal Employment Opportunity Commission

Small Business Administration